

#17



PCT10

RAW SEQUENCE LISTING

DATE: 06/21/2002

PATENT APPLICATION: US/10/019,409A

TIME: 11:35:30

Input Set : A:\10.019.409.txt

Output Set: N:\CRF3\06212002\J019409A.raw

3 <110> APPLICANT: IWAKURA, Masahiro
 5 <120> TITLE OF INVENTION: Sulfur Atom-Free Enzyme Protein
 7 <130> FILE REFERENCE: 4853.0084
 9 <140> CURRENT APPLICATION NUMBER: US 10/019,409A
 C--> 10 <141> CURRENT FILING DATE: 2002-05-13
 12 <150> PRIOR APPLICATION NUMBER: PCT/ JP00/02112
 13 <151> PRIOR FILING DATE: 2000-03-31
 15 <150> PRIOR APPLICATION NUMBER: JP/ 183664
 16 <151> PRIOR FILING DATE: 1999-06-29
 18 <160> NUMBER OF SEQ ID NOS: 10
 20 <170> SOFTWARE: PatentIn version 3.1
 22 <210> SEQ ID NO: 1
 23 <211> LENGTH: 159
 24 <212> TYPE: PRT
 25 <213> ORGANISM: Escherichia coli
 27 <400> SEQUENCE: 1.
 29 Met Ile Ser Leu Ile Ala Ala Leu Ala Val Asp Arg Val Ile Gly Met
 30 1 5 10 15
 33 Glu Asn Ala Met Pro Trp Asn Leu Pro Ala Asp Leu Ala Trp Phe Lys
 34 20 25 30
 37 Arg Asn Thr Leu Asn Lys Pro Val Ile Met Gly Arg His Thr Trp Glu
 38 35 40 45
 41 Ser Ile Gly Arg Pro Leu Pro Gly Arg Lys Asn Ile Ile Leu Ser Ser
 42 50 55 60
 45 Gln Pro Gly Thr Asp Asp Arg Val Thr Trp Val Lys Ser Val Asp Glu
 46 65 70 75 80
 49 Ala Ile Ala Ala Ala Gly Asp Val Pro Glu Ile Met Val Ile Gly Gly
 50 85 90 95
 53 Gly Arg Val Tyr Glu Gln Phe Leu Pro Lys Ala Gln Lys Leu Tyr Leu
 54 100 105 110
 57 Thr His Ile Asp Ala Glu Val Glu Gly Asp Thr His Phe Pro Asp Tyr
 58 115 120 125
 61 Glu Pro Asp Asp Trp Glu Ser Val Phe Ser Glu Phe His Asp Ala Asp
 62 130 135 140
 65 Ala Gln Asn Ser His Ser Tyr Ser Phe Glu Ile Leu Glu Arg Arg
 66 145 150 155
 69 <210> SEQ ID NO: 2
 70 <211> LENGTH: 566
 71 <212> TYPE: DNA
 72 <213> ORGANISM: Escherichia coli
 74 <220> FEATURE:
 75 <221> NAME/KEY: exon
 76 <222> LOCATION: (81)..(557)

ENTERED

RAW SEQUENCE LISTING

DATE: 06/21/2002

PATENT APPLICATION: US/10/019,409A

TIME: 11:35:30

Input Set : A:\10.019.409.txt

Output Set: N:\CRF3\06212002\J019409A.raw

77 <223> OTHER INFORMATION:

80 <400> SEQUENCE: 2

```

81 ggatccttga caattagtta actatttggtt ataatgtatt catgagctta actaactaat      60
83 ccggaagagg aggaacttcc atg atc agt ctg att gcg gcg cta gcg gta gat      113
84           Met Ile Ser Leu Ile Ala Ala Leu Ala Val Asp
85           1           5           10
87 cgc gtt atc ggc atg gaa aac gcc atg cca tgg aac ctg cct gcc gat      161
88 Arg Val Ile Gly Met Glu Asn Ala Met Pro Trp Asn Leu Pro Ala Asp
89           15           20           25
91 ctc gcc tgg ttt aaa cgc aac acc tta aat aaa ccc gtg att atg ggg      209
92 Leu Ala Trp Phe Lys Arg Asn Thr Leu Asn Lys Pro Val Ile Met Gly
93           30           35           40
95 cgc cat acc tgg gaa tca atc ggt agg cct ttg ccc gcc cgc aaa aat      257
96 Arg His Thr Trp Glu Ser Ile Gly Arg Pro Leu Pro Gly Arg Lys Asn
97           45           50           55
99 att atc ctc agc agt caa ccc ggg acc gat gat cgg gtt acc tgg gtt      305
100 Ile Ile Leu Ser Ser Gln Pro Gly Thr Asp Asp Arg Val Thr Trp Val
101 60           65           70           75
103 aaa tcg gtc gac gaa gcc atc gcg gcc gca ggt gac gta cca gaa atc      353
104 Lys Ser Val Asp Glu Ala Ile Ala Ala Ala Gly Asp Val Pro Glu Ile
105           80           85           90
107 atg gtg att ggc ggc gga cgc gtt tat gaa cag ttc ttg cca aaa gcg      401
108 Met Val Ile Gly Gly Gly Arg Val Tyr Glu Gln Phe Leu Pro Lys Ala
109           95           100           105
111 caa aag ctt tat ctg acg cat atc gat gca gaa gtg gaa ggc gac acc      449
112 Gln Lys Leu Tyr Leu Thr His Ile Asp Ala Glu Val Glu Gly Asp Thr
113           110           115           120
115 cat ttt ccg gat tac gag ccg gat gac tgg gaa tcg gta ttc agc gaa      497
116 His Phe Pro Asp Tyr Glu Pro Asp Asp Trp Glu Ser Val Phe Ser Glu
117           125           130           135
119 ttc cac gat gct gat gcg cag aac tcg cat agc tat tcg ttc gaa atc      545
120 Phe His Asp Ala Asp Ala Gln Asn Ser His Ser Tyr Ser Phe Glu Ile
121 140           145           150           155
123 ctc gag cgt cgt taaggatcc      566
124 Leu Glu Arg Arg
128 <210> SEQ ID NO: 3
129 <211> LENGTH: 185
130 <212> TYPE: PRT
131 <213> ORGANISM: Bacillus subtilis
133 <400> SEQUENCE: 3
135 Ala Ser Thr Asp Tyr Trp Gln Asn Trp Thr Asp Gly Gly Gly Ile Val
136 1           5           10           15
139 Asn Ala Val Asn Gly Ser Gly Gly Asn Tyr Ser Val Asn Trp Ser Asn
140           20           25           30
143 Thr Gly Asn Phe Val Val Gly Lys Gly Trp Thr Thr Gly Ser Pro Phe
144           35           40           45
147 Arg Thr Ile Asn Tyr Asn Ala Gly Val Trp Ala Pro Asn Gly Asn Gly
148           50           55           60
151 Tyr Leu Thr Leu Tyr Gly Trp Thr Arg Ser Pro Leu Ile Glu Tyr Tyr

```

RAW SEQUENCE LISTING

DATE: 06/21/2002

PATENT APPLICATION: US/10/019,409A

TIME: 11:35:30

Input Set : A:\10.019.409.txt

Output Set: N:\CRF3\06212002\J019409A.raw

```

152 65          70          75          80
155 Val Val Asp Ser Trp Gly Thr Tyr Arg Pro Thr Gly Thr Tyr Lys Gly
156          85          90          95
159 Thr Val Lys Ser Asp Gly Gly Thr Tyr Asp Ile Tyr Thr Thr Thr Arg
160          100          105          110
163 Tyr Asn Ala Pro Ser Ile Asp Gly Asp Arg Thr Thr Phe Thr Gln Tyr
164          115          120          125
167 Trp Ser Val Arg Gln Ser Lys Arg Pro Thr Gly Ser Asn Ala Thr Ile
168          130          135          140
171 Thr Phe Ser Asn His Val Asn Ala Trp Lys Ser His Gly Met Asn Leu
172 145          150          155          160
175 Gly Ser Asn Trp Ala Tyr Gln Val Met Ala Thr Glu Gly Tyr Gln Ser
176          165          170          175
179 Ser Gly Ser Ser Asn Val Thr Val Trp
180          180          185
183 <210> SEQ ID NO: 4
184 <211> LENGTH: 558
185 <212> TYPE: DNA
186 <213> ORGANISM: Bacillus subtilis
188 <220> FEATURE:
189 <221> NAME/KEY: exon
190 <222> LOCATION: (1)..(555)
191 <223> OTHER INFORMATION:
194 <400> SEQUENCE: 4
195 gct agc aca gac tac tgg caa aat tgg act gat ggg ggc ggt ata gta      48
196 Ala Ser Thr Asp Tyr Trp Gln Asn Trp Thr Asp Gly Gly Gly Ile Val
197 1          5          10          15
199 aac gct gtc aat ggg tct ggc ggg aat tac agt gtt aat tgg tct aat      96
200 Asn Ala Val Asn Gly Ser Gly Gly Asn Tyr Ser Val Asn Trp Ser Asn
201          20          25          30
203 acc gga aat ttt gtt gtt ggt aaa ggt tgg act aca ggt tcg cca ttt      144
204 Thr Gly Asn Phe Val Val Gly Lys Gly Trp Thr Thr Gly Ser Pro Phe
205          35          40          45
207 agg acg ata aac tat aat gcc gga gtt tgg gcg ccg aat ggc aat gga      192
208 Arg Thr Ile Asn Tyr Asn Ala Gly Val Trp Ala Pro Asn Gly Asn Gly
209          50          55          60
211 tat tta act tta tat ggt tgg acg aga tca cct ctc ata gaa tat tat      240
212 Tyr Leu Thr Leu Tyr Gly Trp Thr Arg Ser Pro Leu Ile Glu Tyr Tyr
213 65          70          75          80
215 gta gtg gat tca tgg ggt act tat aga cct act gga acg tat aaa ggt      288
216 Val Val Asp Ser Trp Gly Thr Tyr Arg Pro Thr Gly Thr Tyr Lys Gly
217          85          90          95
219 act gta aaa agt gat ggg ggt aca tat gac ata tat aca act aca cgt      336
220 Thr Val Lys Ser Asp Gly Gly Thr Tyr Asp Ile Tyr Thr Thr Thr Arg
221          100          105          110
223 tat aac gca cct tcc att gat ggc gat cgc act act ttt acg cag tac      384
224 Tyr Asn Ala Pro Ser Ile Asp Gly Asp Arg Thr Thr Phe Thr Gln Tyr
225          115          120          125
227 tgg agt gtt cgc cag tcg aag aga cca acc gga agc aac gct aca atc      432

```

RAW SEQUENCE LISTING

DATE: 06/21/2002

PATENT APPLICATION: US/10/019,409A

TIME: 11:35:30

Input Set : A:\10.019.409.txt

Output Set: N:\CRF3\06212002\J019409A.raw

```

228 Trp Ser Val Arg Gln Ser Lys Arg Pro Thr Gly Ser Asn Ala Thr Ile
229      130      135      140
231 act ttc agc aat cat gtg aac gca tgg aag agc cat gga atg aat ctg      480
232 Thr Phe Ser Asn His Val Asn Ala Trp Lys Ser His Gly Met Asn Leu
233 145      150      155      160
235 ggc agt aat tgg gct tac caa gtc atg gcg aca gaa gga tat caa agt      528
236 Gly Ser Asn Trp Ala Tyr Gln Val Met Ala Thr Glu Gly Tyr Gln Ser
237      165      170      175
239 agt ggc tcg tcg aat gtt acc gta tgg taa      558
240 Ser Gly Ser Ser Asn Val Thr Val Trp
241      180      185
244 <210> SEQ ID NO: 5
245 <211> LENGTH: 159
246 <212> TYPE: PRT
247 <213> ORGANISM: Escherichia coli
249 <400> SEQUENCE: 5
251 Ala Ile Ser Leu Ile Ala Ala Leu Ala Val Asp Arg Val Ile Gly Asn
252 1      5      10      15
255 Glu Asn Ala Leu Pro Trp Asn Leu Pro Ala Asp Leu Ala Trp Phe Lys
256      20      25      30
259 Arg Asn Thr Leu Asn Lys Pro Val Ile Tyr Gly Arg His Thr Trp Glu
260      35      40      45
263 Ser Ile Gly Arg Pro Leu Pro Gly Arg Lys Asn Ile Ile Leu Ser Ser
264      50      55      60
267 Gln Pro Gly Thr Asp Asp Arg Val Thr Trp Val Lys Ser Val Asp Glu
268 65      70      75      80
271 Ala Ile Ala Ala Ala Gly Asp Val Pro Glu Ile Phe Val Ile Gly Gly
272      85      90      95
275 Gly Arg Val Tyr Glu Gln Phe Leu Pro Lys Ala Gln Lys Leu Tyr Leu
276      100      105      110
279 Thr His Ile Asp Ala Glu Val Glu Gly Asp Thr His Phe Pro Asp Tyr
280      115      120      125
283 Glu Pro Asp Asp Trp Glu Ser Val Phe Ser Glu Phe His Asp Ala Asp
284      130      135      140
287 Ala Gln Asn Ser His Ser Tyr Ser Phe Glu Ile Leu Glu Arg Arg
288 145      150      155
291 <210> SEQ ID NO: 6
292 <211> LENGTH: 569
293 <212> TYPE: DNA
294 <213> ORGANISM: Escherichia coli
296 <220> FEATURE:
297 <221> NAME/KEY: exon
298 <222> LOCATION: (81)..(560)
299 <223> OTHER INFORMATION:
302 <400> SEQUENCE: 6
303 ggatccttga caattagtta actatttggtt ataatgtatt catgagctta actaactaat      60
305 ccggaaaagg aggaacttcc atg gca atc agt ctg att gcg gcg cta gcg gta      113
306      Met Ala Ile Ser Leu Ile Ala Ala Leu Ala Val
307      1      5      10

```

RAW SEQUENCE LISTING

DATE: 06/21/2002

PATENT APPLICATION: US/10/019,409A

TIME: 11:35:30

Input Set : A:\10.019.409.txt

Output Set: N:\CRF3\06212002\J019409A.raw

```

309 gat cgc gtt atc ggc aac gaa aac gcc ctc cca tgg aac ctg cct gcc      161
310 Asp Arg Val Ile Gly Asn Glu Asn Ala Leu Pro Trp Asn Leu Pro Ala
311          15          20          25
313 gat ctc gcc tgg ttt aaa cgc aac acc tta aat aaa ccc gtg att tac      209
314 Asp Leu Ala Trp Phe Lys Arg Asn Thr Leu Asn Lys Pro Val Ile Tyr
315          30          35          40
317 ggg cgc cat acc tgg gaa tca atc ggt agg cct ttg ccc ggc cgc aaa      257
318 Gly Arg His Thr Trp Glu Ser Ile Gly Arg Pro Leu Pro Gly Arg Lys
319          45          50          55
321 aat att atc ctc agc agt caa ccc ggg acc gat gat cgg gtt acc tgg      305
322 Asn Ile Ile Leu Ser Ser Gln Pro Gly Thr Asp Asp Arg Val Thr Trp
323 60          65          70          75
325 gtt aaa tcg gtc gac gaa gcc atc gcg gcc gca ggt gac gta cca gaa      353
326 Val Lys Ser Val Asp Glu Ala Ile Ala Ala Ala Gly Asp Val Pro Glu
327          80          85          90
329 atc ttc gtg att ggc ggc gga cgc gtt tat gaa cag ttc ttg cca aaa      401
330 Ile Phe Val Ile Gly Gly Gly Arg Val Tyr Glu Gln Phe Leu Pro Lys
331          95          100          105
333 gcg caa aag ctt tat ctg acg cat atc gat gca gaa gtg gaa ggc gac      449
334 Ala Gln Lys Leu Tyr Leu Thr His Ile Asp Ala Glu Val Glu Gly Asp
335          110          115          120
337 acc cat ttt ccg gat tac gag ccg gat gac tgg gaa tcg gta ttc agc      497
338 Thr His Phe Pro Asp Tyr Glu Pro Asp Asp Trp Glu Ser Val Phe Ser
339          125          130          135
341 gaa ttc cac gat gct gat gcg cag aac tcg cat agc tat tcg ttc gaa      545
342 Glu Phe His Asp Ala Asp Ala Gln Asn Ser His Ser Tyr Ser Phe Glu
343 140          145          150          155
345 atc ctc gag cgt cgt taaggatcc      569
346 Ile Leu Glu Arg Arg
347          160
350 <210> SEQ ID NO: 7
351 <211> LENGTH: 353
352 <212> TYPE: PRT
353 <213> ORGANISM: Artificial Sequence
355 <220> FEATURE:
356 <223> OTHER INFORMATION: fusion protein derived from E. coli and B. subtilis
358 <400> SEQUENCE: 7
360 Ala Ile Ser Leu Ile Ala Ala Leu Ala Val Asp Arg Val Ile Gly Asn
361 1          5          10          15
364 Glu Asn Ala Leu Pro Trp Asn Leu Pro Ala Asp Leu Ala Trp Phe Lys
365          20          25          30
368 Arg Asn Thr Leu Asn Lys Pro Val Ile Tyr Gly Arg His Thr Trp Glu
369          35          40          45
372 Ser Ile Gly Arg Pro Leu Pro Gly Arg Lys Asn Ile Ile Leu Ser Ser
373          50          55          60
376 Gln Pro Gly Thr Asp Asp Arg Val Thr Trp Val Lys Ser Val Asp Glu
377 65          70          75          80
380 Ala Ile Ala Ala Ala Gly Asp Val Pro Glu Ile Phe Val Ile Gly Gly
381          85          90          95

```

VERIFICATION SUMMARY

DATE: 06/21/2002

PATENT APPLICATION: US/10/019,409A

TIME: 11:35:31

Input Set : A:\10.019.409.txt

Output Set: N:\CRF3\06212002\J019409A.raw

L:10 M:271 C: Current Filing Date differs, Replaced Current Filing Date